SURVMETH 745 HW6

Stacey Frank & Chendi Zhao

2/26/2022

# Exercise 9.5

#total budget: 500,000  
#cost: c(1000, 200, 120)  
#delta1: c(.001, .01, .05)  
#delta2: c(.05, .10, .25)  
#pi estimator is used so we're doing srs/srs/srs  
  
#delta1=.001  
Optimum1<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .001, delta2= .05, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
Optimum2<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .001, delta2= .10, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
Optimum3<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .001, delta2= .25, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
  
#delta1=.01  
Optimum4<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .01, delta2= .05, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
Optimum5<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .01, delta2= .10, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
Optimum6<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .01, delta2= .25, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
  
#delta1=.05  
Optimum7<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .05, delta2= .05, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
Optimum8<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .05, delta2= .10, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
Optimum9<- clusOpt3(unit.cost=c(1000, 200, 120), delta1= .05, delta2= .25, unit.rv=2, k1=1, k2=1, tot.cost=500000, cal.sw=1)  
  
delta1<- c(.001,.001,.001, .01,.01,.01, .05, .05, .05)  
delta2<- c(.05, .10, .25,.05, .10, .25,.05, .10, .25)  
m.opt <- c(Optimum1[[10]], Optimum2[[10]], Optimum3[[10]],Optimum4[[10]], Optimum5[[10]], Optimum6[[10]],Optimum7[[10]], Optimum8[[10]], Optimum9[[10]])  
n.opt <- c(Optimum1[[11]], Optimum2[[11]], Optimum3[[11]],Optimum4[[11]], Optimum5[[11]], Optimum6[[11]],Optimum7[[11]], Optimum8[[11]], Optimum9[[11]])  
q.opt <- c(Optimum1[[12]], Optimum2[[12]], Optimum3[[12]],Optimum4[[12]], Optimum5[[12]], Optimum6[[12]],Optimum7[[12]], Optimum8[[12]], Optimum9[[12]])  
CV <- c(Optimum1[[13]], Optimum2[[13]], Optimum3[[13]],Optimum4[[13]], Optimum5[[13]], Optimum6[[13]],Optimum7[[13]], Optimum8[[13]], Optimum9[[13]])  
  
design.options <- data.frame(delta1, delta2, m.opt, n.opt, q.opt, CV)  
  
design.options

## delta1 delta2 m.opt n.opt q.opt CV  
## 1 0.001 0.05 33.7 15.8 5.6 0.0297  
## 2 0.001 0.10 31.5 22.4 3.9 0.0317  
## 3 0.001 0.25 28.5 35.4 2.2 0.0351  
## 4 0.010 0.05 93.0 5.0 5.6 0.0340  
## 5 0.010 0.10 87.7 7.1 3.9 0.0361  
## 6 0.010 0.25 80.2 11.2 2.2 0.0394  
## 7 0.050 0.05 169.1 2.2 5.6 0.0418  
## 8 0.050 0.10 161.2 3.2 3.9 0.0439  
## 9 0.050 0.25 149.6 5.0 2.2 0.0473

The m.opt, n.opt, q.opt, and CV calues for each of the nine combinations of delta1 and delta2 are displayed above. CV is smallest when delta1 and delta2 values are small. Smaller delta1 values are associated with smaller m.opt values, and smaller delta2 values are associated with smaller n.opt values. q.opt gets smaller as delta2 increases (within a given delta1 value).

# Exercise 9.7

## Data

data("MDarea.pop")  
table(MDarea.pop$Hispanic,exclude = NULL)

##   
## 1 2   
## 13202 390795

table(MDarea.pop$Gender,exclude = NULL)

##   
## 1 2   
## 201237 202760

summary(MDarea.pop$Age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 5.00 11.00 10.29 14.00 23.00

## a (three-stage(all srswr))

M <- length(unique(MDarea.pop$TRACT))  
pp.trt1 <- rep(1/M,M)  
MDarea.pop$trtBG <- 10\*MDarea.pop$TRACT + MDarea.pop$BLKGROUP  
res1<-round(rbind(  
 y1=BW3stagePPS(X=MDarea.pop$Hispanic, pp=pp.trt1,  
psuID=MDarea.pop$TRACT, ssuID=MDarea.pop$trtBG),  
y2=BW3stagePPS(X=MDarea.pop$Gender, pp=pp.trt1,  
psuID=MDarea.pop$TRACT, ssuID=MDarea.pop$trtBG),  
y3=BW3stagePPS(X=MDarea.pop$Age, pp=pp.trt1,  
psuID=MDarea.pop$TRACT, ssuID=MDarea.pop$trtBG)),digits = 3)  
rownames(res1)<-c("Hispanic","Gender","Age")  
res1

## B W W2 W3 unit relvar k1 k2 delta1 delta2  
## Hispanic 0.261 0.009 0.261 0.010 0.008 33.089 33.253 0.967 0.963  
## Gender 0.261 0.139 0.257 0.158 0.111 3.607 3.748 0.653 0.619  
## Age 0.242 0.365 0.249 0.406 0.307 1.981 2.135 0.399 0.380

## b (three-stage(ppswr/srs/srs))

pp.trt2 <- table(MDarea.pop$TRACT) / nrow(MDarea.pop)   
MDarea.pop$trtBG <- 10\*MDarea.pop$TRACT + MDarea.pop$BLKGROUP  
res2<-round(rbind(  
 y1=BW3stagePPS(X=MDarea.pop$Hispanic, pp=pp.trt2,  
psuID=MDarea.pop$TRACT, ssuID=MDarea.pop$trtBG),  
y2=BW3stagePPS(X=MDarea.pop$Gender, pp=pp.trt2,  
psuID=MDarea.pop$TRACT, ssuID=MDarea.pop$trtBG),  
y3=BW3stagePPS(X=MDarea.pop$Age, pp=pp.trt2,  
psuID=MDarea.pop$TRACT, ssuID=MDarea.pop$trtBG)),digits = 3)  
  
rownames(res2)<-c("Hispanic","Gender","Age")  
res2

## B W W2 W3 unit relvar k1 k2 delta1 delta2  
## Hispanic 0.000 0.008 0.255 0.009 0.008 1 32.262 0.047 0.966  
## Gender 0.001 0.110 0.249 0.127 0.111 1 3.393 0.007 0.663  
## Age 0.013 0.294 0.248 0.330 0.307 1 1.883 0.043 0.429

## c

For all three variables,when swithcing to a ppswr/srs/srs design to an srs/srs/srs design.,,and decreased slightly and increased slgihtly. However, the value of and became much smaller than before. The is a measure of the homogeneity among the PSU totals. Decrease in this term indicates that the variation within PSUs is much larger than the variation among PSU totals under the ppswr/srs/srs design. This is most obvious in the varible “Hispanis”,“Gender”.